SLFTD: A Subjective Logic based Framework for Truth Discovery

Danchen Zhang[1], Vladimir I. Zadorozhny[1], and Vladimir A. Oleshchuk[2]

[1] School of Computing and Information, University of Pittsburgh, Pittsburgh, USA
[2] Department of Information and Communication Technology, University of Agder, Kristiansand, Norway
{daz45, viz}@pitt.edu; vladimir.oleshchuk@uia.no

Abstract. Finding truth from various conflicting candidate values provided by different data sources is called truth discovery, which is of vital importance in data integration. Several algorithms have been proposed in this area, which usually have similar procedure: iterativly inferring the truth and provider's reliability on providing truth until converge. Therefore, an accurate provider's reliability evaluation is essential. However, no work pays attention to "how reliable this provider continuously providing truth". Therefore, we introduce subjective logic, which can records both (1) the provider's reliability of generating truth, and (2) reliability of provider continuously doing so. Our proposed methods provides a better evaluation for data providers, and based which, truth are discovered more accurately. Our framework can handle both categorical and numerical data, and can identify truth in either a generative or discriminative way. Experiments on two popular real world datasets, Book and Population, validates that our proposed subjective logic based framework can discover truth much more accurately than state-of-art methods.

Keywords: Data Fusion · Truth Discovery · Subjective Logic

1 Introduction

Resolving data conflicts is an important data management problem. For example, for a given flight, different websites may report different departing time; for a given book, different online stores may provide different author lists in their web pages; different studies may report different historical statistics on disease dynamics in the same location. Figuring out true, or the most likely fused data from conflicting values provided by different sources is a challenging task.

Note, that we cannot use a simple voting mechanism, selecting the most popular value as the true value. Take the online book store posting book author list as an example, many online book stores simply copy book author list from another online books store. If we want to find the true author list, voting may be miss-leading. Besides, when several values have same or similar voting count, it is hard to make a decision. In this case, provider's reliability should be taken into consideration. For example, if we know a reliable bookstore that always provides correct author list on its web page, when another less known bookstore provides a different author list, we prefer to trust the former one. Similarly, if we buy a product on eBay, we might trust a more reliable highly ranked store, and buy the product there, even if the price is higher. Therefore, it is very important to accurately evaluate our trustworthiness towards each data provider. Recently, many truth discovery methods have been proposed to deal with this problem [1–4, 6–9, 16, 21, 23–25].

Another challenge in this area is how to select/generate the most reliable data value taking into account provider of providers for each value. Assume there are three providers A, B and C and we assigned high provider to A and B, and low provider to C. When A and B give provide the same data value, but C gives a different value, it is obvious that we should select the first value. However, it may be hard to decide when A, B and C provide three different values. Prior works suggested some basic approaches to solve this problem, such as selecting the value based on maximum average provider reliability.

Though several algorithms haven proposed in this area, but their performance show large difference in different truth discovery scenarios. Also, there is no "one-fits-all" model appearing, different dataset shows different best methods [16]. This motivates us to propose a generalized framework, that can be efficiently adapted to different truth discovery scenarios with less effort but relatively more stable performance.

By introducing subjective logic we can evaluate the data provider's reliability more accurately, and hence discover truth more accurately. We will demonstrate that our approach considerably overcomes the the existing truth discovery methods.

To summarize, our paper has following major contributions:

- In the area of truth discovery, our study is the first to pay attention to "how reliable the provider is able to continuously provide truth". The proposed model can evaluate both (1) the provider's reliability of generating truth, and (2) reliability of provider continuously doing so, providing a more accurate description of data provider's quality.
- Subjective logic is first introduced by us to truth discovery area. Different from commonly used evidence based probabilistic logic, it can perfectly records above mentioned two kinds of reliability for each provider.
- The proposed framework is quite generalized, able to handle both categorical and numerical data, and able to identify truth in either a generative or discriminative way.
- We test our framework on two real-world dataset. The experiments show that compared with state-of-art methods, our framework can improve the truth discovery performance by a large degree.

The rest of the paper is organized as follows: related works are reviewed in Section 2. Background knowledge on Subjective Logic is presented in Section 3. The proposed subjective logic based framework for truth discovery is presented in Section 4. Then we validate the effectiveness of our approach in Section 5. Section 6 concludes this study.

2 Related Works

In this section, we first review the related techniques in the area of truth discovery, and then we elaborate on prior works on subjective logic. Table 1 summarizes the related methods and their features compared to our approach.

Table 1. Summary of state-of-art truth discovery methods.

Method	Provider Count	Provider Reliability	Ability of Entity Discrimination	Value popularity	Value similarity
Voting	√	×	×	×	×
Median	×	×	×	×	×
Average	×	×	×	×	×
Accuracy [1]	√	√	×	×	×
POPAAccuracy [2]	√	√	×	√	×
AccuracySimilarity [1]	√	√	×	×	√
TruthFinder [4]	√	√	×	×	√
AverageLog [21]	√	√	×	×	×
Investment [21]	√	√	×	×	×
PooledInvestment [21]	√	√	×	×	×
SSTF [7]	√	√	×	×	×
2-Estimates [3]	√	√	×	×	×
3-Estimates [3]	√	√	×	×	×
Cosine [3]	√	√	×	×	×
IR based model [3]	√	√	×	×	×
precision/recall [6]	√	√	×	×	×
CRH [23]	√	√	×	×	√
CATD [24]	√	√	×	×	√
GMT [25]	√	√	×	×	√
SLFTD-Dis	√	√	√	×	√
SLFTD-Gen	√	√	√	×	√

In truth discovery area, the simplest mechanism is voting, which does not consider the providers reliability. However, providers reliability assessment is an essential procedure, and many works have been devoted to this area [1, 2, 8, 21, 7, 3, 4]. The most popular are Bayesian based methods [1, 2, 8]. In [1], Dong, et al. proposed to use Accuracy, which is calculated as the probability of each value being correct, and average the confidence of facets provided by the source as the provider trustworthiness. After that, they proposed the concept of AccuracySimilarity, which further considers the similarity of two values. In [2], authors proposed POPAccuarcy, which differs from Accuracy by releasing the assumption that false value probability is uniformly distributed. In [1, 2, 9] they explored the data copying problem, which has not been considered in current work. Another Bayesian method is the TruthFinder, proposed by Yin, et al. [8], which differs from Accuracy by not normalizing the confidence score of each entity.

The second group of methods is based on the web links analysis [21, 7, 3]. In [21], Pasternack, et al., proposed three methods: (1) AverageLog is a transformation of Hub-Authority algorithm, with source trustworthiness being the averaged confidence score of provided values multiplying the log of provided value count; (2) Investment, where the confidence score of the value grows exponentially with the accumulated providers trustworthiness. (3) PooledInvestment, where the confidence score of the value grows linearly. In [7] authors proposed a semi-supervised reliability assessment method, SSTF. It is basically a PageRank method assuming that there is a set of entities having the true value, which will affect the result in the PageRank iteration. [3], proposes 2-Estimates, which is a transformation of Hub-Authority algorithm, whose provider trustworthiness is the average instead of the sum of the vote count. They further proposed 3-Estimates, which additionally considers the values trustworthiness.

Other approaches include IR based methods [3] and precision/recall based techniques [6]. For example, Galland et al. [3] build a vector for each value, with each dimension corresponding to a provider. The reliability of the provider is evaluated as the cosine similarity between provided and selected values. In [6], Pochampally et al., proposed a method measuring the source precision and recall and the correlation information between sources, based on which the value confidence score is computed.

The next group contains CRH [23], CATD [24] and GTM [25]. They are designed to deal with numerical values in a generative manner. With slight modification, they can also be used in categorical dataset.

An accurate provider's reliability evaluation is essential. However, no work pays attention to "how reliable this provider continuously providing truth". Therefore, we introduce subjective logic, which can records both (1) the provider's reliability of generating truth, and (2) reliability of provider continuously doing so. This will be further discussed in next two sections.

Subjective Logic [12, 13] is a powerful decision making tool extending the probabilistic logic by including uncertainty and subjective belief ownership. It is widely used in trust network analysis [14], conditional inference [15], information provider reliability assessment [16], trust management in sensor networks [17], etc. Subjective logic uses subjective opinions to express subjective beliefs about the truth of propositions with degrees of uncertainty. Kane and Browne [18] successfully applied subjective logic to a wireless network environment. In [19], Liu et al, presented a novel reputation computation model to discover and prevent selfish behaviors by combining familiarity values with subjective opinions. To the best of our knowledge, our work is the first one applying it to area of reliable truth discovery.

3 Subjective Logic

In this section, background knowledge of subjective logic is introduced. There are different ways to describe people's opinion towards a statement. Take the classic game "coin toss" for example, people may guess next toss is "head", another one may guess "tail". Binary logic is the simplest way to represent people's opinion, whether 1 or 0 ("head" or "tail"). However, binary logic is usually too simple to describe the full story. Probabilistic logic is the most common way, using an evidence based probability (raining from 0 to 1), to represent people's opinion. For example, after observing the flipping the coin for hundreds of times, people believe the probability of "head" is 0.5, and believe the probability of "tail" is 0.5, too. However, when sample size is too small, the probability is unreliable. In such a situation, subjective logic, proposed by Jøsang [22], can provide more information for the statement.

With subjective logic, an opinion from a person p towards a statement s can be represented by a triple $\omega_s^p = \{t, d, u\}$, with $t, d, u \in [0,1]^3$, and $t + d + u = 1$. In this triple, t means trust, d means distrust, and u means uncertainty. Again take "coin toss" for example, when many enough samples are observed, with half "heads" and half tail, our guess of next toss being "head" can be described as $\{0.49995, 0.49995, 0.0001\}$, using a very small u to describe our ineradicable uncertainty. However, if we only observe the coin flipping for 4 times with half "heads" and half tail, we are still not sure if it is a standard coin. And our guess of next toss being "head" could be described as $\{0.1, 0.1, 0.8\}$, with a high uncertainty u.

Subjective Logic defines a set of logical operations [22], and in this paper we use two of them:

– **Recommendation operation**. Assume two persons, A and B: A has an opinion towards B, and B has an opinion towards a statement s. Then according to B's recommendation, A can generate an opinion towards this statement s. The recommendation operator $\otimes$ is defined as:

  $\omega_s^{AB} = \omega_B^A \otimes \omega_s^B = \{t_s^{AB}, d_s^{AB}, u_s^{AB}\}$, where

  $t_s^{AB} = t_B^A t_s^B$, $d_s^{AB} = t_B^A d_s^B$, and $u_s^{AB} = d_B^A + u_B^A + t_B^A u_s^B$.

– **Consensus operation**. If two persons A and B have opinions towards one statement s, then consensus operator $\oplus$ can be used to combine their opinions. The definition of the consensus operator $\oplus$ is as follows:

  $\omega_s^{A,B} = \omega_s^A \oplus \omega_s^B = \{t_s^{A,B}, d_s^{A,B}, u_s^{A,B}\}$, where

  $t_s^{A,B} = \frac{t_s^A u_s^B + t_s^B u_s^A}{u_s^A + u_s^B - u_s^A u_s^B}$, $d_s^{A,B} = \frac{d_s^A u_s^B + d_s^B u_s^A}{u_s^A + u_s^B - u_s^A u_s^B}$, and $u_s^{A,B} = \frac{u_s^A u_s^B}{u_s^A + u_s^B - u_s^A u_s^B}$.

With the recommendation and consensus operations, people can merge their opinions towards an unknown entity by other people's or information-provider's recommendation. For example, person A wants to know whether a new movie is worthy to watch, as shown in Figure 1. He searches online and finds a blog saying the movie is absolutely the best movie of the year. However, his friend B told him that he watched the movie yesterday, and it is quite disappointing. We can assume A's opinion towards B (i.e., the statement "B is trustful") is

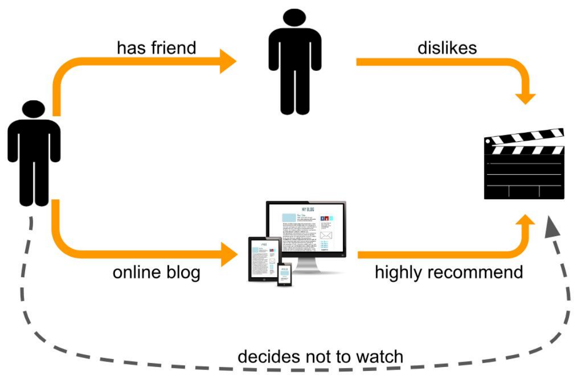

Fig. 1. The procedure of deciding whether to watch a movie.

{0.8, 0.1, 0.1}, while B's impression towards the movie (i.e., the statement "the movie worth money and time") is {0.5, 0.4, 0.1}. Then, according to B's recommendation, A will have an opinion towards the movie $\{0.4, 0.32, 0.28\}$. Then, let's calculate A's opinion with blog's recommendation. A does not fully trust the online blogs, since sometimes the blog uses overpraised words to attract people consuming, and we assume A's opinion about the blog is $\{0.5, 0.2, 0.3\}$. The blog holds an opinion $\{1, 0, 0\}$ towards the movie. According to the blog's recommendation, A's opinion towards the movie is $\{0.5, 0, 0.5\}$. After combining two opinions, A has a final impression towards the movie, $\{0.53, 0.25, 0.22\}$, and decides not to watch it.

4 Subjective logic based framework for truth discovery (SLFTD)

4.1 Problem Formulation

Consider a dataset that contains a set of entities $E = \{e_1, e_2, ..., e_n\}$, and a set of data providers $P = \{p_1, p_2, ..., p_m\}$, the value of entity e_i provided by provider p_j is named as v_{ij}, constructing the value set V. Different providers may provide different value for same entity, truth discovery aims to find the true value for each entity. Such a dataset can be represented as a matrix shown in Table 2. In this matrix, each row corresponds to an entity, each column corresponds to a provider, and the cell represents a value that the provider assigns to the corresponding entity. If a provider does not provide a value for an entity, the cell value is empty.

4.2 Proposed framework

Current methods usually have an iterative two-step procedure to obtain estimated true value: (1) Inferring truth of each entity. Given the provider's reliability, the value with largest support from providers should be true. (2) Assigning

Table 2. The dataset is represented by a matrix, with n entities and m providers.

	$\mathbf{p_1}$	$\mathbf{p_2}$	$\mathbf{p_3}$	...	$\mathbf{p_j}$	...	$\mathbf{p_m}$
$\mathbf{e_1}$	v_{11}	v_{12}	v_{13}	...	v_{1j}	...	v_{1m}
$\mathbf{e_2}$	v_{21}	v_{22}	v_{23}	...	v_{2j}	...	v_{2m}
$\mathbf{e_3}$	v_{31}	v_{32}	v_{33}	...	v_{3j}	...	v_{3m}
...	...	...	...	...	...	...	...
$\mathbf{e_i}$	v_{i1}	v_{i2}	v_{i3}	...	v_{ij}	...	v_{im}
...	...	...	...	...	...	...	...
$\mathbf{e_n}$	v_{n1}	v_{n2}	v_{n3}	...	v_{nj}	...	v_{nm}

reliability to providers. If a provider frequently provides true values, it should be assigned with a higher score. After data being converged, the inferred truth is outputted.

Different from past methods, we propose to evaluate the provider's reliability score and entity's discrimination score in an iterative way, and then for each provider, subjective logic based opinions are constructed based on the converged scores, as shown in Figure 2 Step 1. After that, the true value are inferred based on the fused opinions in either a discriminative manner or a generative manner, as shown in Figure 2 Step 2.

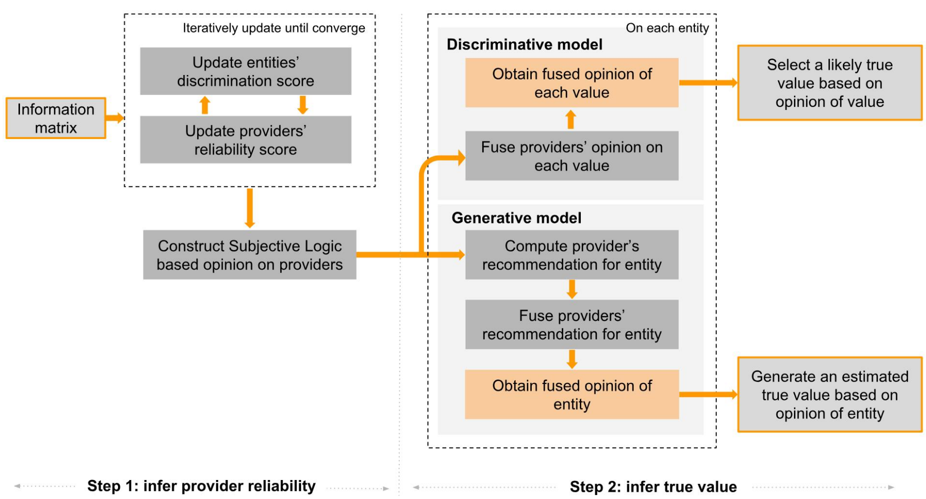

Fig. 2. System flow of subjective logic based truth discovery framework.

4.3 Accurately infer providers' reliability

In the weight iteration procedure of Step 1, together with provider's reliability score, we propose to infer the entities' discrimination score instead of the value's score of being true. Our consideration is that, given an entity, if most of the candidate values from reliable providers are very similar/close to each other, the true value is very likely to be one of them or very close to them. However, given another entity, if candidate values from the reliable providers vary largely, even for a human, it is hard to infer the true value, and the inferred score for each value is less unconvincing. When evaluate a provider's reliability, the first entity can provide a more convincing evidence than the second entity. Further, improper truth inference may lead to chain-reacting effect in the following iterative rounds. Based on such consideration, we propose to iterativly evaluate provider's reliability score and entity's discrimination score.

In this study, the degree, to which the algorithm can infer the true value of the entity in an undisputed convincing manner, is defined as the **entity discrimination ability**. The entity whose majority candidate values are from reliable providers and are very similar/close to each other should be given a higher score. Thus discrimination score of entity E_i is defined as:

$$Disc(E_i) = \frac{\sum_{P_j, P_l \in P_{E_i}} Rel(P_j)Rel(P_l)Imp(V_{il} \rightarrow V_{ij})}{\sum_{P_j, P_l \in P_{E_i}} Rel(P_j)Rel(P_l)} \tag{1}$$

where P_{E_i} is the set of providers that gives value on entity E_i; $Rel(P_j)$ is the reliability score of provider P_j, which will be described later. Please notice that $Imp(V_{il} \rightarrow V_{ij})$ reflects the implication from $V_{i.}$ to V_{ij}, introduced from [8]. It is a value reflecting to what degree V_{ij} is (partially) true if V_{il} is correct. In this study, Imp range is set from 0 to 1, with 0 means no such implication, and 1 means fully support V_{ij} is true. At the first iteration, all provider have equal weights; and after each round, normalization is conducted with weights summing up to 1.

When measure the **providers' reliability**, instead of no difference, our framework pays more attention to providers' performance on entities with higher discrimination score. Given such entities, if the values from this provider obtain lots of implications from other values of same entity, this provider reliability should be boosted; otherwise, should be lower down. Such impact from entities with low discrimination score should be relatively discounted. Thus reliability score of provider P_j is defined as:

$$Rel(P_j) = \frac{\sum_{E_i \in E_{P_j}} Disc(E_i)Imp(V_{i.} \rightarrow V_{ij})}{\sum_{E_i \in E_{P_j}} Disc(E_i)} \tag{2}$$

where E_{P_j} is the set of entities to whom P_j gives value; and $V_{i.}$ is a value set, consisting of all candidate values of entity E_i.

Also, normalization is conducted in the end of each iteration. This iterative procedure will continue until all scores converge.

After entity discrimination score and provider reliability score being converged, subjective logic based opinions towards each provider is computed before inferring truth. Agreeing with past studies, true value should be the one with the highest provider support. However, a single reliability score cannot describe the whole story for the provider, because how reliable is this reliability score stays unknown. For example, given two providers having same reliability score 0.95, the first provider provides values for 50 entities with high discrimination score, while the second provider only provides values for 1 entity with low discrimination score. Intuitively, reliability score of the first provider should be more convincing than that of the second provider.

Thus, two kinds of provider reliability should be considered: (1) reliability of generating the true value; (2) reliability of continuously doing so. The above proposed reliability score can be used to describe the first kind of reliability, and we propose a new concept, **certainty**, to describe the second one. Certainty of provider P_j is defined as:

$$Certainty(P_j) = \sum_{E_i \in E_{P_j}} Disc(E_i) \tag{3}$$

To fit such data, we propose to introduce subjective logic to truth discovery, which can record an evidenced-based score, and also the confidence on it. Each triple pinion consists of trust, distrust, and uncertainty. We proposed to record the provider's reliability of generating the true value in trust; and to record the reliability of continuously doing with uncertainty. In this way, our trustworthiness to the provider in two aspects can better described. The proposed algorithm $Algo$'s subjective logic based opinion towards the P_j's reliability, $\omega_{P_j}^{Algo} = \{t_{P_j}^{Algo}, d_{P_j}^{Algo}, u_{P_j}^{Algo}\}$, is defined as:

$$t_{P_j}^{Algo} = (1 - u_{P_j}^{Algo})Rel(P_j) \tag{4}$$

$$d_{P_j}^{Algo} = 1 - t_{P_j}^{Algo} - u_{P_j}^{Algo} \tag{5}$$

$$u_{P_j}^{Algo} = \gamma(1 - Certainty(P_j)) + \alpha \tag{6}$$

where α describe people's fundamental uncertainty, since even given by enough evidence, people can still be skeptical. γ is a parameter to limit the certainty to a certain range. Both parameters range from 0 to 1. In this way, provider's reliability can be accurately described.

4.4 Infer the true value with subjective logic based opinions

When inferring the true value in the Step 2, there are two strategies: (1) discriminative model, selecting a most likely true value from all candidates; (2) generative model, generating a true value which may does not appear in the dataset. The former can works on both categorical and numerical data, while the latter can only work on the numerical data.

Infer truth in generative manner. This model only fits numerical data, i.e., $V_{ij} \in R$. To utilize the generated subjective logic based opinions, for each entity, we propose to generate the opinion towards the statement "true value of entity is the max candidate value". If such an opinion can be obtained, higher trust means truth is close to the max candidate value; otherwise, truth is close to the min candidate value.

First, on each entity E_i, we normalize all the candidate values in the following manner:

$$V_{ij}^{'} = \frac{V_{ij} - min(V_{i.})}{max(V_{i.}) - min(V_{i.})}, \tag{7}$$

so that $V_{ij}^{'} \in [0, 1]$. In this way, statement "true value of entity is the max candidate value" is mapped to "true value of entity in the normalized space is 1". Thereby, given provider E_i, the provider P_j's opinion towards the statement can be defined as:

$$\omega_{truth(E_i)=1}^{P_j} = \{(1-\beta)V_{ij}^{'}, 1-(1-\beta)V_{ij}^{'} - \beta, \beta\}, \tag{8}$$

where β also describe people's fundamental uncertainty, similar to α.

Second, the provider can recommend his opinion of the entity's truth to the algorithm. As mentioned in Section 3, recommendation operation can help people know the statement according to their acquaintances. Thus, algorithm's opinion towards truth of E_i by P_j's recommendation is defined as:

$$\omega_{truth(E_i)=1}^{Algo,P_j} = \omega_{P_j}^{Algo} \otimes \omega_{truth(E_i)=1}^{P_j}. \tag{9}$$

Entity E_i has a set of candidate values from several providers $\{P_j, ..., P_k\}$, and the algorithm should have a summarized opinion based on all recommendations. Consensus operation in Section 3 can help fuse several opinions towards one statement together. The algorithm's final opinion towards truth of E_i is defined as:

$$\omega_{truth(E_i)=1}^{Algo,P_j,...,P_k} = \omega_{truth(E_i)=1}^{Algo,P_j} \oplus ... \oplus \omega_{truth(E_i)=1}^{Algo,P_k}. \tag{10}$$

In the final fused opinion, the trust reflects the true value of E_i in the normalized space, and final step is to map it to the original numerical space by:

$$V_{ij}^{true} = t_{truth(E_i)=1}^{Algo,P_j,...,P_k}(max(V_{i.}) - min(V_{i.})) + min(V_{i.}). \tag{11}$$

Infer truth in discriminative manner. In this model, for each candidate value in each entity, we propose to generate a fused subjective logic based opinion, and then for each entity, select the candidate with highest trust as the truth.

Skipping the recommendation procedure, given a provider P_j, the algorithm's opinion towards a value V_{ij} is directly copied from its opinion towards P_j, which is defined as:

$$\omega_{V_{ij}}^{Algo,P_j} = \{t_{P_j}^{Algo}, d_{P_j}^{Algo}, u_{P_j}^{Algo}\}. \tag{12}$$

If a value is provided by several providers $\{P_j, ..., P_k\}$, consensus operation is used to fuse opinions together. Thus we have algorithm's final opinion towards a value V_{ij}:

$$\omega_{V_{ij}}^{Algo,P_j,...,P_k} = \omega_{V_{ij}}^{Algo,P_j} \oplus ... \oplus \omega_{V_{ij}}^{Algo,P_k}. \tag{13}$$

5 Experiments

In this section, we evaluate our proposed framework on two popular real word datasets, one being categorical another being numerical. Also, its performance is compared with state-of-art algorithms.

5.1 Proposed methods and baselines

SLFTD-Gen. This is the proposed method SLFTD generating the true value in a generative way.

SLFTD-Dis. This is the proposed method SLFTD selecting the true value from existing candidates in a discriminative manner.

Voting. The candidate with max amount of providers is true data. If several candidates receive same voting, randomly pick one.

Median; Average. The median and average of all candidate values is predicted as true.

Sums; Average.log; Investment; PooledInvestment; TruthFinder; Accuracy; AccuracySim. These seven discriminative methods have similar main idea, iterativly update each value's score and provider's reliability, only in different computing manners. First five methods appear in [21], and last two are proposed in [9]. TruthFinder and AccuracySim considers the similarity between candidate values, while other methods do not.

CATD; CRH; GTM. These three models are designed as generative model for numerical data, but can be adapted to categorical data as a discriminative model with slight modification. Each iteration, with evaluated provider's reliability, they try to generate/select estimated true value of each entity to minimize the difference between "estimated true matrix" and the observed input matrix [24, 23, 25]. Additionally, CATD is designed to smoothly predict truth on the long tail data with chi-squared distribution. The extra merits of first two methods is the lack of parameters.

5.2 Finding true book author list

Dataset: Book. It is a popular categorical dataset in truth discovery area, composed by Luna Xin Dong [3]. Its data describes that for each book, online bookstores post author list in their web pages, but some data is wrong. It contains the information on ISBN, book name, authors, online bookstore name for 1265 books. Totally, there are 894 bookstores and they generate 26,494 author lists.

We have two testing data. First one is the gold testing dataset published by Luna Dong, consisting of 100 books. The second testing dataset is composed of 161 book, containing the first 100 books and other 61 books. The 61 books are selected because different methods appearing in our experiments generates different true data. Thus it is more challenging than the first one. Similar to Luna Dong, we call it silver testing dataset. For both testing data, the true author list are manually assigned by people reading the cover page of the book. In our experiments, we will report the accuracy of each method on both testing dataset.

Since we do not have access to the pre-processed dataset used in previous works, we do the data cleaning by ourselves, and the clean data is posted online [4]. In the dataset, most stores separate the names by ";", but many others use ",". We manually recognize those stores and change them to names separated by ";". Then following [1], middle names are removed. Our dataset is cleaner compared to the data used in prior works, since, as we will see below, the voting results in our case is 82%, while past studies showed only 71%.

Settings. The implication appeared in Equation 1, is defined as $Imp(V_{il} \rightarrow V_{ij}) = \frac{\#|V_{il} \bigcap V_{ij}|}{\#|V_{ij}|}$, where $\#|V_{il}|$ is the amount of elements in V_{il}; the implication appeared in Equation 2 is defined as $Imp(V_{i\cdot} \rightarrow V_{ij})$ is defined as $\frac{\sum_{P_l, P_j \in P_{E_i}} Imp(V_{il} \rightarrow V_{ij})}{\sum_{P_l, P_j \in P_{E_i}} 1}$.

Following past studies, the parameters of all methods are set with optimal performance on the testing data. In TruthFinder, λ is set to be 0.4. In AccuracySim, λ is set to be 0.9. For the proposed method SLFTD-Dis, both α and γ are set to be 0.2.

Results. Precision of eleven methods are shown in Table 3, which is sorted by the performance on silver testing data. We can see that our proposed method SLFTD-Dis has the best performance on both testing data. Further, SLFTD-Dis increases precision by 3.3% compared with the second best method AccuracySim on the golden testing data; and is better than the second best method PooledInvestment by 6.7% on silver testing data. In addition, it seems that discriminative models have a much better performance than the CRH and CATD, which are modified to adapt this task. Also, methods (SLFTD, TruthFinder, AccuracySim)

[3] http://lunadong.com/fusionDataSets.htm
[4] http://crystal.exp.sis.pitt.edu:8080/daz45/

Table 3. Precision of eleven methods on true book author list finding task. Best results are in bolder.

Method	Golden Testing	Silver Testing
SLFTD-Dis	**0.94**	**0.776**
PooledInvestment	0.87	0.7275
TruthFinder	0.86	0.708
AccuracySim	0.91	0.689
Accuracy	0.89	0.689
Investment	0.79	0.634
Average.log	0.82	0.621
Voting	0.80	0.621
Sums	0.74	0.553
CRH	0.4	0.304
CATD	0.4	0.304

that utilize the similarity/implication between values also shows a better performance than those who does not use.

5.3 Finding true population of the city

Dataset: Population. In this study, we pick the dataset Population, proposed in [21], to validate our proposed framework. This is a numerical dataset, a sample of Wikipedia edit history of city population. When the data was released in 2010, there were 44,761 tuples from 4,107 data providers. The version used in [25, 23, 24] contains 43,071 tuples. When we download it in 2019, it contains 51,761 tuples from 4,264 data providers on 40,583 cities. The testing data stays same, consisting of 308 randomly collected cities manually labeled with true population. Therefore, the experiment results differs from the results from past papers. We pre-process the dataset in a same way as [25, 23, 24]: (1) One provider may provide several population to same city, only the latest one is kept. (2) if a city only have one candidate value (from one or several providers), its data is removed. (3) Outliers on each city are removed in the same way as [25] with TruthFinder. After pre-processing, compared with 4,119 tuples on 1,148 cities from 2,415 providers are left and methods are evaluated on 274 cities [25, 23, 24], in our experiment dataset, 5,731 tuples on 1,814 cities from 2,467 providers are left, and methods are evaluated on 280 cities, which can be accessed in the same URL[5].

Settings. The implication appeared in Equation 1, is defined as: $Imp(V_{il} \rightarrow V_{ij}) = 1 - \frac{|V_{ij} - V_{il}|}{max(V_{i.}) - min(V_{i.})}$; the implication appeared in Equation 2 is defined as $Imp(V_{i.} \rightarrow V_{ij}) = 1 - \frac{|V_{ij} - avg(V_{i.})|}{max(V_{i.}) - min(V_{i.})}$.

[5] http://crystal.exp.sis.pitt.edu:8080/daz45/

Following [25, 23, 24], three evaluation metrics are selected: MAE, RMSE, and Error Rate. In terms of Error Rate, "error" appears when the predicted truth is smaller or larger than the ground truth by 10%.

Similarly, following past studies, the parameters of all methods are set based on optimal performance on the testing data. In TruthFinder, λ is set to be 0.3. In terms of GTM, we have two set of parameters, the first being ($\alpha = 10, \beta = 10, \mu_0 = 0, \sigma_0^2 = 1$) suggested by [25], and the second being ($\alpha = 4, \beta = 1, \mu_0 = 0, \sigma_0^2 = 1$), which has best performance in our experiment. For the proposed method SLFTD-Dis, γ is set to be 0.2, while α is set to be 0. Finally, for SLFTD-Gen, γ is set to be 0.2, while α is set to be 0.8, and β is set to be 0.6.

Table 4. Precision of eleven methods on true book author list finding task. First group shows the performance of four methods without removing outliers; second group shows the performance on the data without outliers, and predict truth in a generative manner; methods in third group also works on data without outliers, and predict truth in a discriminative way. Best results are in bolder; second best is labeled with *.

Methods	MAE	RMSE	Error Rate
TruthFinder	3586.936	12545.950	0.129*
Voting	3982.425	13713.190	0.157
Median	11319.105	124967.792	0.193
Average	6.51E+15	9.90E+16	0.636
SLFTD-Gen	**2821.356**	**8057.032**	**0.125**
GTM - best parameters	4185.617	11079.278	0.214
Average	4380.647	11105.030	0.232
CRH	4442.703	11341.750	0.214
Median	3811.934	11863.048	0.175
GTM - parameters in [25]	4326.951	11992.095	0.168
CATD	3991.858	12772.711	0.161
SLFTD-Dis	3385.243*	10732.963*	0.129*
Voting	3674.471	12412.074	0.150
TruthFinder	3610.543	12552.167	0.129*
Average.log	4303.643	13368.106	0.157
Sums	4410.111	13486.296	0.157
Investment	4435.793	14152.204	0.182
PooledInvestment	4549.221	14162.000	0.193

Results. All methods' performance is shown in Table 4. We can see that the proposed method SLFTD-Gen gives best performance on all three metrics. Additionally, SLFTD-Dis gives the second best on all three metrics. We can also see following findings: (1) It is reasonable to use predictions from TruthFinder as priors to remove outliers, consistent with findings from [25]. Naive methods, especially Average, gives a much worse performance. (2) Second group (generative models) usually have a relatively smaller RMSE and and a higher Error Rate than the third group (discriminative models), indicating that either the "correct cases" whose distance is smaller than 10% from truth in the second group are more accurate than that in the third group, or the "error cases" in third group are farther from truth than that of second group. (3) Also, the lower Error Rate in third group means that true value usually appears in the candidate value set. Also, we tried to run TruthFinder after outlier removed, but it does not provide further improvement, and even declined a little bit in terms of MAE and RMSE.

6 Conclusion

In this study, we proposed a subjective logic based framework for the truth discovery, which can predict truth either in a discriminative way or a generative manner. Subjective logic is introduced because it is able to describe the provider's reliability more accurately, and it also provide sufficient opinion operators to manipulate opinions from different sources. Experiments on two real world datasets validates that our proposed framework can discover truth more accurately than state-of-art methods.

In the next step, we will conduct experiments to explicitly check the evaluated reliability of providers with synthetic data. Also, the framework can be designed with other structures, which we would like to try. Additionally, we would like to use our proposed methods in more areas, such as fake news detection, and cancer driver gene discovery.

References

1. Dong, Xin Luna, Laure Berti-Equille, and Divesh Srivastava. "Integrating conflicting data: the role of source dependence." Proceedings of the VLDB Endowment 2.1 (2009): 550-561.
2. Dong, Xin Luna, Barna Saha, and Divesh Srivastava. "Less is more: Selecting sources wisely for integration." Proceedings of the VLDB Endowment. Vol. 6. No. 2. VLDB Endowment, 2012.
3. Galland, Alban, et al. "Corroborating information from disagreeing views." Proceedings of the third ACM international conference on Web search and data mining. ACM, 2010.
4. Li, Xian, et al. "Truth finding on the deep web: Is the problem solved?." Proceedings of the VLDB Endowment. Vol. 6. No. 2. VLDB Endowment, 2012.
5. Pasternack, Jeff, and Dan Roth. "Knowing what to believe (when you already know something)." Proceedings of the 23rd International Conference on Computational Linguistics. Association for Computational Linguistics, 2010.
6. Pochampally, Ravali, et al. "Fusing data with correlations." Proceedings of the 2014 ACM SIGMOD international conference on Management of data. ACM, 2014.
7. Yin, Xiaoxin, and Wenzhao Tan. "Semi-supervised truth discovery." Proceedings of the 20th international conference on World wide web. ACM, 2011.
8. Yin, Xiaoxin, Jiawei Han, and S. Yu Philip. "Truth discovery with multiple conflicting information providers on the web." IEEE Transactions on Knowledge and Data Engineering 20.6 (2008): 796-808.

9. Dong, Xin Luna, and Divesh Srivastava. "Big data integration." Data Engineering (ICDE), 2013 IEEE 29th International Conference on. IEEE, 2013.
10. Oleshchuk, Vladimir, and Vladimir Zadorozhny. "Trust-aware query processing in data intensive sensor networks." sensorcomm. IEEE, 2007.
11. Karlik, Bekir, and A. Vehbi Olgac. "Performance analysis of various activation functions in generalized MLP architectures of neural networks." International Journal of Artificial Intelligence and Expert Systems 1.4 (2011): 111-122.
12. Jsang, Audun. Subjective logic. Springer International Publishing Switzerland, 2016.
13. Jsang, Audun. "Artificial reasoning with subjective logic." Proceedings of the second Australian workshop on commonsense reasoning. Vol. 48. Perth:[sn], 1997.
14. Jsang, Audun, Ross Hayward, and Simon Pope. "Trust network analysis with subjective logic." Proceedings of the 29th Australasian Computer Science Conference-Volume 48. Australian Computer Society, Inc., 2006.
15. Josang, Audun. "Conditional reasoning with subjective logic." Journal of Multiple-Valued Logic and Soft Computing 15.1 (2008): 5-38.
16. Pelechrinis, Konstantinos, et al. "Automatic evaluation of information provider reliability and expertise." World Wide Web 18.1 (2015): 33-72.
17. Oleshchuk, Vladimir, and Vladimir Zadorozhny. "Trust-aware query processing in data intensive sensor networks." sensorcomm. IEEE, 2007.
18. Kane, Kevin, and James C. Browne. "Using uncertainty in reputation methods to enforce cooperation in ad-hoc networks." Proceedings of the 5th ACM workshop on Wireless security. ACM, 2006.
19. Liu, Yining, et al. "A novel reputation computation model based on subjective logic for mobile ad hoc networks." Future Generation Computer Systems 27.5 (2011): 547-554.
20. Lawrence, Michael S., et al. "Mutational heterogeneity in cancer and the search for new cancer-associated genes." Nature 499.7457 (2013): 214.
21. Pasternack, Jeff, and Dan Roth. "Knowing what to believe (when you already know something)." Proceedings of the 23rd International Conference on Computational Linguistics. Association for Computational Linguistics, 2010.
22. Jsang, Audun. Subjective logic. Heidelberg: Springer, 2016.
23. Li, Qi, et al. "Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation." Proceedings of the 2014 ACM SIGMOD international conference on Management of data. ACM, 2014.
24. Li, Qi, et al. "A confidence-aware approach for truth discovery on long-tail data." Proceedings of the VLDB Endowment 8.4 (2014): 425-436.
25. Zhao, Bo, and Jiawei Han. "A probabilistic model for estimating real-valued truth from conflicting sources." Proc. of QDB (2012).